Bat echolocation behavior from highspeed 3D video

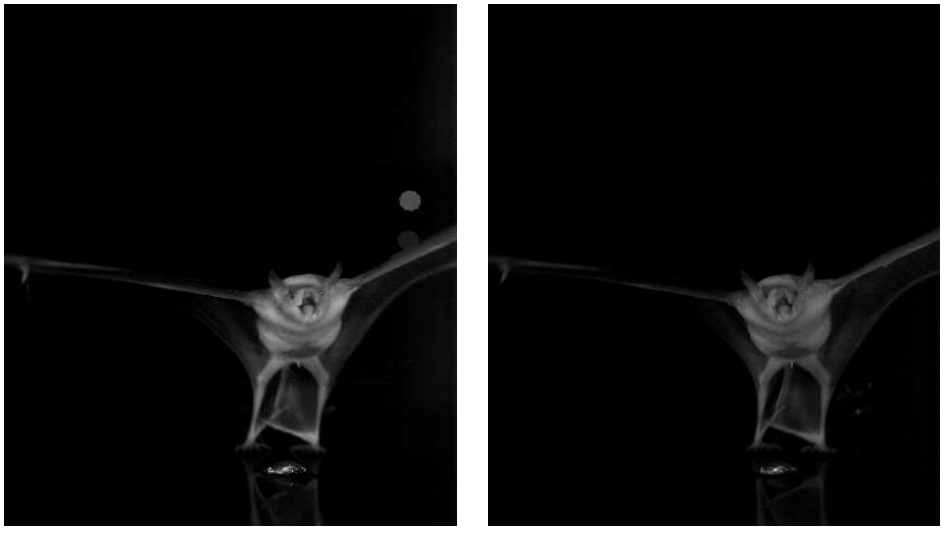
R. Fisher, Y. Xiao
University of Edinburgh

1 Introduction

As part of the ChiRoPing research project[1] we have been observing the 3D position and orientation of several different species of bats, as well as making measurements of different aspects of their bodies. The long-term goal is to understand what the bats are doing during echolocation, identification and tracking of prey, so that robot sonar sensors with a comparable level of performance can be developed. We claim that registered highspeed 3D video can expose behavior not observable using conventional or even high speed video sensors.

There is previous monocular research on insect and animal tracking (*e.g.* [2]). There is also starting to be a stream of 3D video capture for face analysis and synthesis, *e.g.* [4]), and for some medical analysis *e.g.* [3]). We are not aware of any previous research on high-speed 3D shape tracking and analysis.

The main data for this research comes from a bespoke commercial 500 frame/second stereo vision system (Dimensional Imaging), which consists of a pair of low light sensitive stereo cameras that are synchronized for stereo acquisition. The data streams into a 2 second ring buffer, which is stopped by the operator after an interesting event occurs (usually prey captured by the bat). The data is then uploaded to disk for stereo computation offline. While slightly more than 1000 frames of data are acquired, most of the species move through a 30-40 cm capture zone sufficiently quickly so that only 30-100 frames of the 1000 are useful. Each frame of data has about 1.2 million depth measurements, with a registered left infrared intensity image and a corresponding right intensity image. Analysis of the camera calibration suggests about 0.1 mm RMS noise level on individual 3D point positions when using automatically matched static points, and about 0.2-1.0 mm rms on moving points (depending on the speed and direction of motion). More details of the sensor and its performance can be seen in [1]. An example of a left and right image of a *Noctilio leporinus* is here, with the prey at its feet.



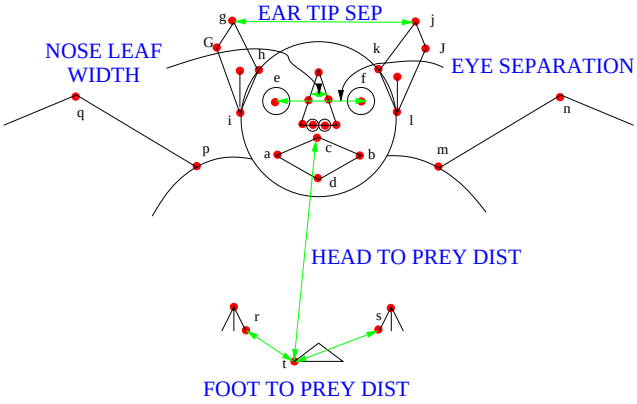
This paper describes some of the analysis of bat capture data from March 2009, at the Smithsonian Tropical Research Center, on Panama's Barro Colorado Island. Data from three species were captured, but

[1] EC IST programme, STREP project 215370, ICT Challenge 2: "Cognitive Systems, Interaction, Robotics". See `http://www.chiroping.org/`

only results from *N. leporinus* are discussed here. The data comes from wild bats that have been brought into a flight cage, where they can perform reasonably natural prey hunting behavior. In the case of *N. leporinus*, this is trawling for small fish and other small prey on a water surface in front of the cameras.

2 Measurements and Methodology

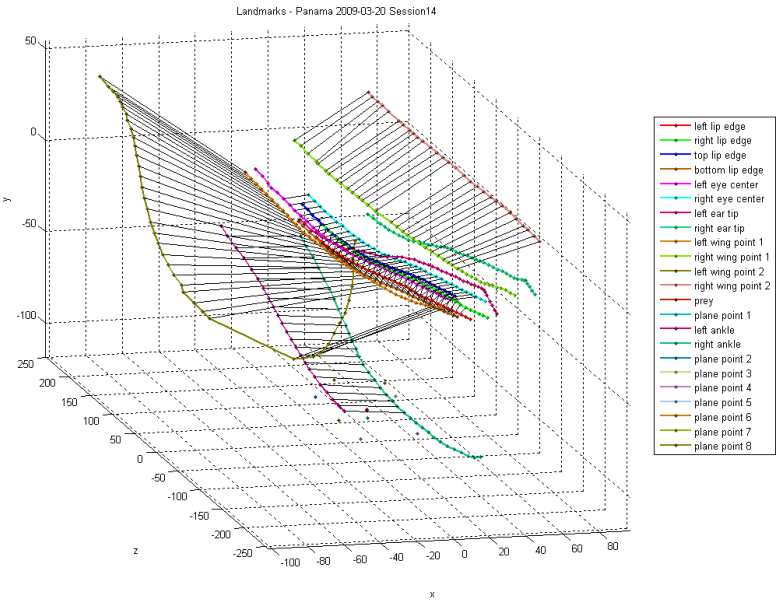
Because of low contrast and motion blur, it was not possible to accurately and automatically extract corresponding points from the stereo image data. So, the corresponding points are marked by hand, at least initially. Area correlation tracking from frame-to-frame works with some well-defined features. Other features needed manual marking in each frame. After this manual step, 3D reconstruction, point tracking and property measurement are automatic through the dataset. The camera system is calibrated using Dimensional Imaging's software, which gives the camera projection matrices. Given the corresponding left and right image feature points, our software back-projects using the camera projection matrices and intersects the view rays to estimate a 3D position for each feature point. As the semi-manual annotation has linked the 2D point positions over time, this automatically transfers to the 3D point positions. In the figure below, the dots mark the feature points that we extract (when visible). The figure shows some of the 3D measurements that we compute using the estimated 3D point positions.



Some of the 30+ measurements that are computed for each video frame are defined formally as follows. Let $\vec{n}_d = (n_{dx}, n_{dy}, n_{dz})' = \vec{D}/ \mid\mid \vec{D} \mid\mid$, where $\vec{D} = (\vec{c} - \frac{1}{2}(\vec{e} + \vec{f}))$ be the head direction vector (invert if necessary so $n_{dz} > 0$). Let $\vec{n}_m = \frac{(\vec{e}-\vec{f})\times(\vec{f}-\vec{c})}{||(\vec{e}-\vec{f})\times(\vec{f}-\vec{c})||}$ be the upward facing unit normal vector of the $\vec{e} - \vec{f} - \vec{c}$ plane (invert if necessary so $n_{my} > 0$).

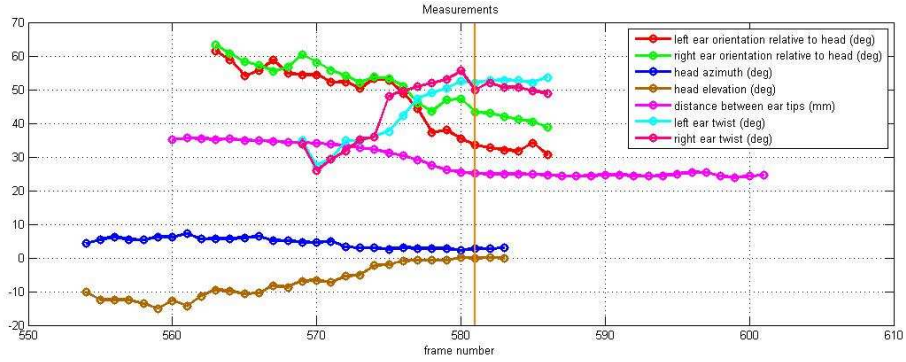
1. Eye separation $\mid\mid \vec{f} - \vec{e} \mid\mid$. Mainly for consistency checking.
2. Mouth height $||\vec{c} - \vec{d}||$ and width $||\vec{b} - \vec{a}||$.
3. Head azimuth and elevation. Azimuth is $atan(\frac{n_{dx}}{n_{dz}})$, where 0 corresponds to heading towards the camera and positive is to the right of the camera. Elevation is $atan(\frac{n_{dy}}{\sqrt{n_{dx}^2+n_{dz}^2}})$, where 0 corresponds to heading towards the camera and positive is above the camera.
4. Foot-to-prey distance: $||\vec{r} - \vec{t}||$ or $||\vec{s} - \vec{t}||$, depending on which foot is used to capture the prey. If the mouth is used, then $||\vec{c} - \vec{t}||$.
5. Wing angle: $acos(\frac{\vec{q}-\vec{p}}{||\vec{q}-\vec{p}||} \cdot \frac{\vec{n}-\vec{m}}{||\vec{n}-\vec{m}||})$
6. Distance between ear tips: $\vec{g} - \vec{j}$.
7. Ear orientation relative to head: $acos(\frac{\vec{g}-\vec{h}}{||\vec{g}-\vec{h}||} \cdot \vec{n}_m)$ and $acos(\frac{\vec{j}-\vec{k}}{||\vec{j}-\vec{k}||} \cdot \vec{n}_m)$.
8. Ear twist: $acos(\frac{\vec{G}-\vec{h}}{||\vec{G}-\vec{h}||} \cdot \frac{\vec{e}-\vec{f}}{||\vec{e}-\vec{f}||})$ and $acos(\frac{\vec{J}-\vec{k}}{||\vec{J}-\vec{k}||} \cdot \frac{\vec{f}-\vec{e}}{||\vec{f}-\vec{e}||})$.

An example of the trajectories mapped by some of the selected points is next shown. You can see the 3D point positions linked together through time.

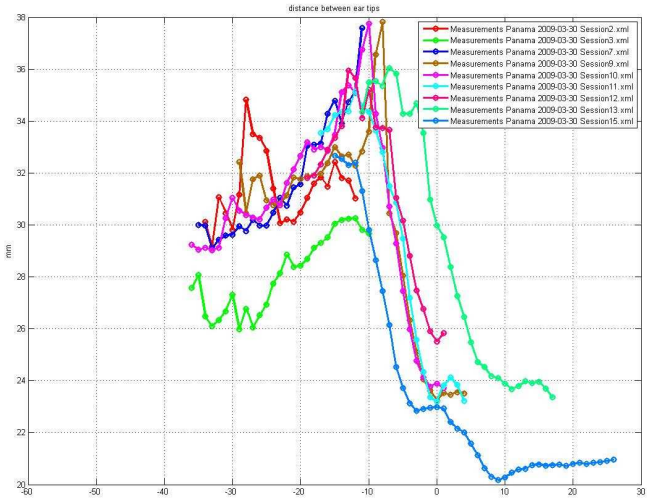


3 Results

From the 3D measurements observed over time, some behaviorally interesting results can be computed. An example of the 3D measurements for a *N. leporinus* bat from a single sequence is shown below. The horizontal axis is the frame number and the vertical tan bar is the frame at which the bat's foot contacts the prey. We can see that the bat is flying at a constant forward azimuth (blue) and tilts its head upward (brown elevation). While this is happening, the ear orientations raise (red and green angles with the head 'vertical' reduce) and the ears also twist outward (purple and cyan angles relative to the eye separation). The distance between the ear tips (pink) also reduces. What we can simultaneously see in the original video data is the bat twisting from a gliding position (frame 555) to an upright position where it can use its feet to grab the prey (frame 582), while simultaneously raising its ears.



Based on 3D measurement plots like these and also visual observations, bat zoologists and bio-acoustic specialists have hypothesized that at the start the bat's ears are focusing on the prey, whereas after committing to prey capture, the ears are raised to concentrate on the environment after it commits to a capture. While it is hard to know the bat's intentions, one can use the 3D data to at least see that this behavior is consistent. The next plot shows the ear separation for 9 different observations of the same *N. leporinus* (Panama 20/3/09). The data are aligned by shifting the data so that all estimated prey contact times are at time 0. The plot shows the distance between the ear tips, which is a reliable measurement and highly correlated with the ear elevation. What we can see consistently is the separation increasing to a peak about 10 frames (20 msec) before capture, and then falling.



4 Discussion

Without the high-speed video, it would not have been possible to observe the ear movements, as the events took place over 30-40 frames, which is 60-80 msec. This corresponds to 2-3 frames in a normal video sensor, which would have captured much more blurred data as the bat is moving at about 4 m/s during this time. Without the 3D measurements, we would not have been able to detect and measure the dipping movement before the ear-lift, as this was not obvious in the individual camera's recording. Thus both 3D and high speed make important contributions to understanding what the bat is doing.

A major limitation of this project is the need to manually annotate the feature points. Some automation is used through inter-frame image cross correlation, however this drifts even on well defined points requiring some re-initialization. We are considering how to exploit the 3D geometric generic bat head models we have from micro-CT scanning. In theory, they can be deformably matched to the noisy 3D data, where the deformations take account of both individual and plastic deformation differences. However, so far we find that it is even difficult for human annotation, given the low contrast motion blurred images.

Although this paper is about the image analysis, the project is also recording acoustic data from up to 50 microphones simultaneously to determine when the bat emits its ultrasonic pulses and what is the energy distribution of those pulses. This will also give cues as to what the bat is doing.

This abstract has concentrated on only 1 species, but we are looking at similar phenomena over 4 species, with on the order of 200 capture sequences. The morphology and behavior of each species is slightly different, so we hope to discover species-specific strategies (*i.e.* morphological, acoustic and behavioral) for prey detection and capture, and then adapt these strategies to advanced robot sonar systems.

References

[1] Y. Xiao, R. B. Fisher, M. Oscar, "Performance Characterization of a High-Speed Stereovision Sensor for Acquisition of Time-Varying 3D Shapes", Machine Vision and Applications, in press.

[2] Y. Bachalany, F. Cabestaing, C. Vieren, S. Ambellouis, R. Olberg, Tracking Dragonflies in Image Sequences, Proc. Workshop on Visual observation and analysis of animal and insect behavior, Tampa, 2008.

[3] Benton, L., Nebel, J.-C., Study of the breathing pattern based on 4D data collected by a dynamic 3D body scanner, Proc. 7th Numerisation 3D/Scanning 2002 Congress, Paris, France, (2002).

[4] Ypsilos, I.A., Hilton, A., Turkmani, A., Jackson, P.,. Speech driven face synthesis from 3D video, IEEE Symposium on 3D Data Processing, Visualisation and Transmission, pp.58-65 (2004).